The Short Musculoskeletal Function Assessment

de Graaf, Max W.; Reininga, Inge H. F.; Wendt, Klaus W.; Heineman, Erik; El Moumni, Mostafa

Published in:
Clinical Rehabilitation

DOI:
10.1177/0269215519828152

Publication date:
2019

Article

CLINICAL REHABILITATION

Clinical Rehabilitation
2019, Vol. 33(5) 923–935

DOI: 10.1177/0269215519828152
journals.sagepub.com/home/cre
SAGE

The Short Musculoskeletal Function Assessment: a study of the reliability, construct validity and responsiveness in patients sustaining trauma

Max W de Graaf[1], Inge HF Reininga[1], Klaus W Wendt[1], Erik Heineman[2] and Mostafa El Moumni[1]

Abstract

Objective: To assess test–retest reliability, construct validity and responsiveness of the Dutch Short Musculoskeletal Function Assessment (SMFA-NL) in patients who sustained acute physical trauma.
Design: A longitudinal cohort study.
Setting: A level 1 trauma center in The Netherlands.
Subjects: Patients who required hospital admission after sustaining an acute physical trauma.
Intervention: Patients completed the SMFA-NL at six weeks, eight weeks and six months post-injury.
Main measure: The measures used were The Dutch Short Musculoskeletal Function Assessment. Test–retest reliability (between six and eight weeks post-injury) using intraclass correlation coefficients, the smallest detectable change and Bland and Altman plots. Construct validity (six weeks post-injury) and responsiveness (between six weeks and six months post-injury) were evaluated using the hypothesis testing method.
Results: A total of 248 patients (mean age: 46.5, SD: 13.4) participated, 145 patients completed the retest questionnaires (eight weeks) and 160 patients completed the responsiveness questionnaires (six months). The intraclass correlation coefficients indicated good to excellent reliability on all subscales (0.80 to 0.98). The smallest detectable change was 17.4 for the *Upper Extremity Dysfunction* subscale, 11.0 for the *Lower Extremity Dysfunction* subscales, 13.9 for the *Problems with Daily Activities* subscale and 16.5 for the *Mental and Emotional Problems* subscale. At group level, the smallest detectable change ranged from 1.48 to 1.96. A total of 86% of the construct validity hypotheses and 79% of the responsiveness hypotheses were confirmed.

[1]Department of Trauma Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[2]Department of Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Corresponding author:
Max W de Graaf, Department of Trauma Surgery, University Medical Center Groningen, University of Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.
Email: m.w.de.graaf@umcg.nl

Conclusion: This study showed that the SMFA-NL has good to excellent reliability, sufficient construct validity and is able to detect change in physical function over time.

Keywords
Functional status, clinimetric, patient-reported outcome, rehabilitation, trauma

Date received: 28 March 2018; accepted: 11 January 2019

Introduction

Patient-reported outcome measures have become increasingly important to evaluate functional outcome after trauma. Numerous patient-reported outcome measures have been developed for this purpose, yet most are disease- or body-region-specific, hence not suitable to assess heterogeneous samples such as patients who sustained various kinds of injuries. In 1999, Swiontkowski et al.[1] introduced the Short Musculoskeletal Function Assessment (SMFA), a patient-reported outcome measure that can be used to gauge physical functioning in patients with a broad range of musculoskeletal conditions. The SMFA was designed as an instrument that is not too specific, nor overly general and is considered suitable for heterogeneous samples such as patients with a broad range of traumatic injuries.[2,3]

The SMFA originally consisted of two indices: the Function Index (34 items) and the Bother Index (12 items).[1] The Function Index was considered to be a relatively strict measure of functional limitations, while the Bother Index indicated the "amount" of bother due to the functional limitations. The SMFA has been cross-culturally adapted in various languages, including Dutch (SMFA-NL).[4–9]

The two indices were originally reported to be valid, reliable and responsive,[10,11] although recently the Function and Bother indices were shown to have insufficient structural validity in Dutch trauma patients using the SMFA-NL.[12] The findings indicated that the two indices are not a valid representation of the latent construct physical functioning. Only a four-subscale configuration consisting of the subscales *Upper Extremity Dysfunction, Lower Extremity Dysfunction, Problems with Daily Activities* and *Mental and Emotional Problems* demonstrated sufficient structural validity.[12]

Although the four-subscale structure showed sufficient structural validity in a broad range of trauma patients, additional clinimetric properties (reliability, construct validity and responsiveness) have not yet been evaluated. Evaluation of these properties is required to justify usage in clinical and research settings. The aim of this study was therefore to evaluate test–retest reliability, construct validity and responsiveness of the four subscales of the SMFA-NL in patients with a broad range of traumatic injuries.

Methods

Study design and recruitment of patients

A longitudinal cohort study design was used. Patients were recruited between October 2012 and March 2016 at University Medical Center Groningen, a level-1 trauma center in the Netherlands. The methods employed in this study have been reviewed by the local Institutional Review Board, which waived further need for approval (METc2012.104). Patients consented to participate in this study. The study was conducted in compliance with the principles outlined in the Declaration of Helsinki on ethical principles for medical research involving human subjects.

The inclusion criterion was patients admitted to the hospital due to acute traumatic injuries. Exclusion criteria were age between 18 and 65, patients who could not read or write Dutch, injuries that resulted in severe neurological deficits, pathological fractures and patients with severe psychiatric conditions (such as active psychosis, bipolar disorders, major depressive episodes).

Patients were asked to complete six questionnaires (described below) at six weeks post-injury

and to complete the SMFA-NL for a second time after a two-week interval. Patients received the six questionnaires again at six months post-injury. Standard questionnaires were used. Patients received the questionnaires on paper or electronically and non-responders were reminded once.

Outcome measures

Short Musculoskeletal Function Assessment. The SMFA-NL contains 46 items that can be divided into four subscales: *Upper Extremity Dysfunction, Lower Extremity Dysfunction, Problems with Daily Activities* and *Mental and Emotional Problems*.[9,12] All items are scored on a 5-point Likert-type scale. The items of each of the SMFA-NL subscales can be summed up and divided by the maximum score to create subscales, each ranging from 0 to 100, where 0 represents best possible function.

Health Utilities Index 3. The Health Utilities Index 3 is a validated 15-item generic health questionnaire that can be used to assess specific Health-Related Quality of Life and specific health domains including Ambulation, Dexterity, Emotion and Pain.[13] The Health-Related Quality of Life score (Multi Attribute Score) ranges from 0 to 1, with a score of 1 as best. The standard "past one-week" version was used.[13] The Health Utilities Index 3 has been recommended and shown valid in a wide range of conditions, including patients with acute traumatic injuries.[13–19] The Health Utilities Index 3 is available in several languages, including Dutch.[13]

EuroQoL-5 Dimensions. The EuroQoL-5 Dimensions (EQ-5D) questionnaire is a generic instrument that can be used to assess health status and Health-Related Quality of Life.[20] The EQ-5D consists of five items scored on a 3-point scale and from which a single index score can be calculated. The score ranges from 0 (representing death) to 100 (representing optimal health).[21,22] The EQ-5D has been recommended and shown valid for assessing health status and Health-Related Quality of Life in trauma patients.[14,15,17,23–25] It is available in over 180 languages, including Dutch.[26]

Disabilities of Arm, Shoulder and Hand. The Disabilities of Arm, Shoulder and Hand is a body region-specific questionnaire that can be used to assess upper extremity dysfunction.[27] It consists of 30 items that are scored on a 5-point Likert-type scale, from which a total score can be calculated. The score ranges from 0 to 100, where 0 represents best possible function. The Disabilities of Arm, Shoulder and Hand has been cross-culturally adapted in various languages, including Dutch, and has been validated in patients with upper extremity injuries.[17,28–32]

Lower Extremity Functional Scale. The Lower Extremity Functional Scale is a body region-specific questionnaire that can be used to assess lower extremity function.[33] It consists of 20 items scored on a 5-point Likert-type scale. Items are summed to a total score ranging from 0 to 80. A score of 80 represents the best possible function. The Lower Extremity Functional Scale has been cross-culturally adapted in Dutch and shown to be valid for assessing lower extremity function in patients with traumatic injuries of the lower extremity.[34–38]

Numeric Pain Rating Scale. The 11-point Numeric Pain Rating Scale is a valid and frequently used unidimensional measurement instrument to assess pain in adults.[39,40] Scores ranged from 0 to 10 in discrete numbers, where 0 indicated no pain at all and 10 represented the worst imaginable pain.

Global Rating of Effect. Global Rating of Effect questions were used to verify whether no clinical change had occurred in the test–retest interval. The Global Rating of Effect questions were specified for all four subscales of the SMFA-NL, with five answer options ranging from "much improved" to "much deteriorated."

Procedures

Clinimetric properties were assessed in accordance to the COSMIN guidelines.[41] Test–retest reliability[42] of the SMFA-NL was evaluated using the six and eight weeks post-injury measurements.

Construct validity[42] was assessed with the six weeks post-injury data. A total of 50 hypotheses were predefined in terms of expected direction and expected magnitude of correlations of the SMFA-NL with the following patient-reported outcome measures and clinical parameters (Table 3 and Supplemental Appendix 1). Outcome measures used were Health Utilities Index 3; EQ-5D; Disabilities of Arm, Shoulder and Hand; Lower Extremity Functional Scale; and Numeric Pain Rating Scale. Clinical parameters were Injury Severity Score, anatomical injury region, surgical complications reported within 30 days of the injury and hospital length of stay. Injury severity scores were obtained from the Dutch Trauma Registry and institutional patient registry.[43] Surgical complications were obtained from the institutional complication registry.

Responsiveness[42] was assessed using the six weeks and six months post-injury data. Hypotheses were predefined for the expected correlation between changes in scores on the SMFA-NL and changes in scores on the Health Utilities Index 3; EQ-5D; Disabilities of Arm, Shoulder and Hand; Lower Extremity Functional Scale; and Numeric Pain Rating Scale questionnaires (Table 4). Additional hypotheses were predefined for discriminative capacity between groups of patients based on anatomical injury region or whether a surgical complication was reported within six months post-injury (Supplemental Appendix 2).

Statistical analysis

To assess clinimetric properties, a sample size of at least 50 patients is considered minimal and 100 patients preferable.[44] Anticipating a 40% loss to follow-up and 10%–15% of all patients missing one or more items in any of the returned questionnaires, we aimed to include at least 200 patients.

Test–retest reliability was evaluated using the intraclass correlation coefficient (ICC 2,1) for absolute agreement and was based on a two-way random effects model.[45] Only patients who scored the Global Rating of Effect question of the specific subscale as "not changed" were included in the test–retest analysis. Intraclass correlation coefficients $\geqslant 0.70$ were considered an indication of good reliability and values $\geqslant 0.90$ an indication of excellent reliability.[44]

Measurement error was evaluated with the standard error of measurement for absolute agreement, smallest detectable change and limits of agreement in Bland and Altman plots.[45] The smallest detectable change was calculated at the individual and group level.[46] In the Bland and Altman plots, the difference in scores between the test–retest measurements were plotted against the mean score of the test–retest measurements.[47] Limits of agreement were calculated as mean test–retest difference $\pm$ 1.96 $SD_{difference}$. One-sample *t*-tests were used for all subscales to determine whether the difference between the test and retest measurement was different from zero. A significant difference was considered evidence of systematic bias.[47] Univariable linear regression analyses were used to investigate proportional bias: the effect of the mean test–retest scores on the test–retest difference scores. Regression coefficients that were statistically different from zero were considered to be an indication of proportional bias.[47]

Construct validity was considered sufficient when at least 75% of the predefined hypotheses were confirmed.[44] The predefined hypotheses were tested using Pearson correlation coefficients for continuous variables, and mean differences were calculated to assess differences between specific groups of patients. Confirmation or rejection of the hypotheses was based on the magnitude of the correlation coefficient or mean difference, rather than *P*-values.[41] A correlation coefficient <0.3 was considered low, between 0.3 and 0.59 moderate, and $\geqslant 0.6$ high.

The data of the six weeks post-injury measurement was assessed for floor and ceiling effects. Floor or ceiling effects occur when patients score the absolute maximum or minimum score on a measurement instrument. When $\geqslant 15\%$ of the measurements were either the minimum or maximum score, they were regarded as a floor or ceiling effect, respectively.[48] Patients without upper or lower extremity injuries may be expected to report the best possible score on the *Upper* or *Lower Extremity Dysfunction* subscales, respectively. Hence, floor and ceiling effects on the *Upper* and

Lower extremity subscales were analyzed in patients who had an upper or lower extremity injury, respectively. The entire study sample was used to analyze floor and ceiling effects for the *Problems with Daily Activities* and *Mental and Emotional Problems* subscales.

A total of 42 predefined hypotheses (Table 4, Supplemental Appendix 2) on responsiveness of the SMFA-NL were tested using Pearson correlation coefficients. For responsiveness, both measurements may carry measurement error; therefore, correlation coefficients <0.25 were considered low, between 0.25 and 0.49 moderate and $\geqslant 0.5$ high.[49] Responsiveness was considered sufficient when at least 75% of the predefined hypotheses were confirmed.[44]

Results

A total of 248 patients completed the questionnaires at six weeks post-injury. The response rate was 64%. The general characteristics of the study sample are shown in Table 1. Most patients were treated surgically and the lower extremity was the most common anatomical region of injury (Table 1). Several patients did not disclose marital status or educational level (Table 1). In total, 145 patients completed both the test and retest questionnaire. A total of 160 patients completed both the six weeks and six months post-injury questionnaires.

Clinimetric properties

The intraclass correlation coefficients of the *Upper Extremity Dysfunction* and *Mental and Emotional Problems* subscales indicated good reliability (Table 2). The *Lower Extremity Dysfunction* and *Problems with Daily Activities* subscales demonstrated excellent reliability (Table 2). The standard error of measurement and smallest detectable change are shown in Table 2. Least measurement error was demonstrated for the *Lower Extremity Dysfunction* subscale, indicating best precision among the four subscales. The *Upper Extremity Dysfunction Subscale* demonstrated most measurement error. Bland and Altman plots do not show an upward or downward trend for any of the subscales (Figure 1). The measurements were equally spread above and below the 0 line for all subscales. The limits of agreement were smallest for the *Lower Extremity Dysfunction* and *Problems with Daily Activities* subscales and widest for the *Upper Extremity Dysfunction* subscale. The mean test–retest differences of the subscales were not significantly different from zero for the *Upper Extremity*

Table 1. General characteristics of the study sample.

General characteristics	*N* (%)
Gender ($n=248$)	
Male	148 (60)
Female	100 (40)
Age ($n=248$)	46.5 (13.4)[a]
Marital status ($n=218$)	
Single	75 (33)
With partner	144 (67)
Educational level ($n=206$)	
Elementary school	3 (1)
High school	70 (31)
College	70 (31)
Bachelor's degree or higher	81 (36)
Other	4 (1)
Injuries ($n=678$)	
Head and neck	40 (6)
Face	30 (4)
Thorax	62 (9)
Abdomen	25 (4)
Spine	98 (14)
Upper extremity	155 (23)
Lower extremity and pelvic bones	214 (32)
Skin[b]/other	54 (8)
Injury Severity Score ($n=248$)	
All patients	4 (1–42)[c]
Major trauma (ISS $\geqslant$ 16)	35 (14)
Treatment ($n=248$)	
Conservative treatment	43 (17)
Surgery[d]	205 (83)
Surgical complication within 30 days ($n=248$)	36 (15)

ISS: Injury Severity Score.
[a]Presented as mean (SD).
[b]Superficial injuries (abrasion, contusion, lacerations, regardless of anatomical region).
[c]Presented as median (range).
[d]Requiring surgery for at least one of the injuries.

Table 2. Reliability, measurement error, and systematic bias and proportional bias.

	Subscales of the SMFA-NL			
	UED	LED	PDA	MEP
$Mean_{test}$ (SD)	8.9 (19.0)	23.9 (26.8)	53.7 (29.2)	19.8 (13.4)
$Mean_{retest}$ (SD)	8.8 (18.9)	22.4 (26.5)	52.0 (28.8)	19.8 (13.1)
Reliability and measurement error				
$ICC(2,1)_{agr}$	0.89	0.98	0.97	0.80
(95% CI)	(0.84–0.93)	(0.96–0.99)	(0.95–0.98)	(0.69–0.87)
SEM_{agr}	6.28	3.97	5.03	5.95
SDC_{ind}	17.4	11.0	13.9	16.5
SDC_{gr}	1.93	1.48	1.95	1.96
Systematic bias				
Mean difference	−0.10	−1.55	−1.76	0.00
P-value	0.9	0.04	0.07	1.0
Proportional bias				
β	−0.01	−0.06	−0.56	−0.03
P-value	0.9	0.7	0.7	0.8

UED: Upper Extremity Dysfunction; LED: Lower Extremity Dysfunction; PDA: Problems with Daily Activities; MEP: Mental and Emotional Problems; Test: six weeks post-injury; retest: eight weeks post-injury; $ICC(2,1)_{agr}$: intraclass correlation coefficient for agreement using a two-way random effects model; SEM_{agr}: standard error of measurement for agreement; SDC_{ind}: smallest detectable change at the individual level; SDC_{gr}: smallest detectable change at the group level; β: standardized regression coefficient; SMFA-NL: Short Musculoskeletal Function Assessment.

Dysfunction, Problems with Daily Activities and *Mental and Emotional Problems* subscales (Table 2), indicating there was no evidence for systematic bias. Systematic bias was observed for the *Lower Extremity Dysfunction* subscale (Table 2). None of the regression coefficients were significantly different from zero, indicating there was no evidence of proportional bias (Table 2).

The correlation of the SMFA-NL subscales with other patient-reported outcome measures and clinical parameters is shown in Table 3. In total, 43 of the 50 (86%) pre-specified hypotheses were confirmed. All correlation coefficients that were expected to be high were confirmed as such. All but one of the hypotheses expected to have a low correlation were confirmed. Five out of the six hypotheses on discriminative validity were confirmed (Supplemental Appendix 1). Patients who suffered a surgical complication within 30 days scored 14.5 points higher on the *Problems with Daily Activities* subscale (Supplemental Appendix 1). Patients with an upper extremity injury scored 20.5 points higher on the *Upper Extremity Dysfunction* subscale than patients without an upper extremity injury. Patients with a lower extremity injury scored 31.4 points higher on the *Lower Extremity Dysfunction* subscale than those without a lower extremity injury.

A floor effect was observed for the *Upper Extremity Dysfunction* subscale: among patients with an upper extremity injury, 23 patients (20%) scored the lowest possible score. In this group, 16 (70%) patients had a fractured clavicle or scapula, or had a small injury to the hand. Other subscales did not show floor or ceiling effects.

The expected Pearson correlation coefficients of changes in scores between the six weeks and six months post-injury measurement on the SMFA-NL and changes in score on other patient-reported outcome measures are shown in Table 4 and Supplemental Appendix 2. Of the 43 predefined hypotheses, 34 (79%) were confirmed. The *Upper Extremity Dysfunction* subscale showed high correlations with the Health Utilities Index 3 Dexterity subscale and a low correlation with Health Utilities Index 3 Ambulation. The *Upper Extremity Dysfunction* subscale showed a correlation of 0.37

Table 3. Construct validity hypotheses for the SMFA-NL with other instruments and parameters.

	Upper Extremity Dysfunction	Lower Extremity Dysfunction	Problems with Daily Activities	Mental and Emotional Problems
EQ-5D Index	E: moderate (–) O: –0.18	E: high (–) O: **–0.71**	E: high (–) O: **–0.76**	E: moderate (–) O: **–0.49**
HUI3 Multi Attribute Score	E: moderate (–) O: **–0.32**	E: high (–) O: **–0.64**	E: high (–) O: **–0.73**	E: moderate (–) O: **–0.57**
HUI3 Emotion	E: low O: **–0.06**	E: low O: **–0.14**	E: low (–) O: **–0.20**	E: moderate (–) O: **–0.36**
HUI3 Pain	E: moderate (–) O: –0.12	E: moderate (–) O: **–0.36**	E: moderate (–) O: **–0.44**	E: moderate (–) O: **–0.34**
HUI3 Ambulation	E: low O: **0.07**	E: high (–) O: **–0.83**	E: high (–) O: **–0.66**	E: low O: **–0.29**
HUI3 Dexterity	E: high (–) O: **–0.79**	E: low O: **0.13**	E: moderate (–) O: –0.17	E: low O: **0.14**
DASH	E: high (+) O: **0.61**	E: low O: 0.46	E: high (+) O: **0.69**	E: moderate (+) O: **0.46**
LEFS	E: low O: **–0.02**	E: high (–) O: **–0.88**	E: high(–) O: **–0.83**	E: moderate (–) O: **–0.40**
Numeric Pain Rating Scale	E: moderate (+) O: 0.24	E: moderate (+) O: 0.24	E: moderate (+) O: **0.36**	E: moderate (+) O: **0.43**
ISS	E: low O: **–0.07**	E: low O: **0.15**	E: low O: **0.17**	E: low O: **0.20**
Hospital length of stay	E: low O: **0.12**	E: moderate (+) O: **0.32**	E: moderate (+) O: **0.35**	E: low O: **0.24**

E: expected direction and magnitude of predefined correlations of the SMFA-NL subscales with other instruments and parameters; high: $r \geq 0.6$; moderate: $0.3 \leq r < 0.6$; low: $r < 0.3$; (+) or (–): expected direction of correlation; O: observed correlation; HUI3: Health Utilities Index Mark 3; DASH: Disabilities of the Arm Shoulder and Hand; LEFS: Lower Extremity Functional Scale; ISS: Injury Severity Score; SMFA-NL: Short Musculoskeletal Function Assessment.
Expected low correlations not assigned a direction since it was hypothesized that the correlation coefficient would be close to zero. Confirmed hypotheses are shown in bold.

with the Disabilities of Arm, Shoulder and Hand. The *Lower Extremity Dysfunction* subscale showed high correlations with the EQ-5D; Health Utilities Index 3 Ambulation subscale; Disabilities of Arm, Shoulder and Hand; and Lower Extremity Functional Scale. The *Problems with Daily Activities* and *Upper* and *Lower Extremity Dysfunction* subscales showed a low correlation with Health Utilities Index 3 Emotion. The change in *Problems with Daily Activities* score showed high correlations with the EQ-5D; Health Utilities Index 3 Multi Attribute Score; Health Utilities Index 3 Ambulation; Disabilities of Arm, Shoulder and Hand; and Lower Extremity Functional Scale. Hypotheses on the *Mental and Emotional Problems* subscale were confirmed least. Raw scores and change in scores of the six weeks to six months interval are shown in Table 5.

Discussion

This study demonstrated that the SMFA-NL has good to excellent test–retest reliability, sufficient construct validity and responsiveness to assess physical function in patients who sustained trauma. Due to floor effects, the clinical usability of the *Upper Extremity Dysfunction* subscale may be limited.

To justify the use of the SMFA-NL in clinical practice or in applied research, it is important to establish its clinimetric measurement properties in concordance with the COSMIN criteria.[41] The reliability and validity of the SMFA-NL enable assessment of

Table 4. Responsiveness hypotheses for the SMFA-NL with other instruments.

	Upper Extremity Dysfunction	Lower Extremity Dysfunction	Problems with Daily Activities	Mental and Emotional Problems
EQ-5D Index	E: moderate (–) O: –0.18	E: high (–) O: **–0.61**	E: high (–) O: **–0.62**	E: moderate (–) O: **–0.34**
HUI3 Multi Attribute score	E: moderate (–) O: **–0.25**	E: moderate (–) O: **–0.43**	E: high (–) O: –0.47	E: moderate (–) O: **–0.27**
HUI3 Emotion	E: low O: **0.01**	E: low O: **–0.03**	E: low O: **–0.11**	E: high (–) O: –0.13
HUI3 Pain	E: low O: **–0.11**	E: low O: **0.17**	E: moderate (–) O: **–0.31**	E: moderate (–) O: –0.20
HUI3 Ambulation	E: low O: **–0.04**	E: high (–) O: **–0.67**	E: high (–) O: –0.44	E: low O: **–0.09**
HUI3 Dexterity	E: high (–) O: **–0.68**	E: low O: **0.03**	E: moderate (–) O: –0.11	E: low O: **0.05**
DASH	E: high (+) O: 0.37	E: low O: 0.57	E: high (+) O: **0.71**	E: moderate (+) O: **0.37**
LEFS	E: low O: **–0.11**	E: high (–) O: **–0.65**	E: high (–) O: **–0.64**	E: moderate (–) P: **–0.26**
Numeric Pain Rating Scale	E: low O: **0.12**	E: low O: **0.21**	E: moderate (+) O: **0.37**	E: moderate (+) O: **0.28**

E: expected direction and magnitude of predefined correlations of the SMFA-NL subscales with other instruments and parameters; high: $r \geqslant 0.5$; moderate: $0.25 \leqslant r < 0.5$; low: $r < 0.25$; (+) or (–): expected direction of correlation; HUI3: Health Utilities Index Mark 3; DASH: Disabilities of the Arm Shoulder and Hand; LEFS: Lower Extremity Functional Scale; ISS: Injury Severity Score; SMFA-NL: Short Musculoskeletal Function Assessment; O: observed correlation; EQ-5D: EuroQoL-5 Dimensions.
Expected low correlations were not assigned a direction since it was hypothesized that the correlation coefficient would be close to zero. Confirmed hypotheses are shown in bold.
Hypotheses were formulated as follows: The correlation of change in … score with change in … score is expected to be ….
For example, the correlation of change in EQ-5D index score with change in SMFA-NL Problems with Daily Activities score is expected to be high.

physical function at a single point in time. The responsiveness of the SMFA-NL allows evaluation of (recovery of) physical function over time. The SMFA-NL can be applied in patients with a broad range of injuries, ranging from wounds to major trauma.

Previous studies have assessed the clinimetric properties of the SMFA-NL in trauma patients, using a slightly modified version where double-barreled items were split.[50,51] Van Son et al.[51] assessed the clinimetric properties in patients with isolated unilateral lower extremity fractures and upper extremity fractures. They reported sufficient reliability, construct validity and responsiveness, yet in that study two different three-subscale structures were used to calculate scores: one set of subscales for patients with upper extremity fractures and another set of subscales for patients with lower extremity fractures. This complicates the scoring, especially in patients who suffered fractures of both the upper and lower extremities. Van Delft-Schreurs et al.[50] assessed the clinimetric properties in a sample that only contained major trauma patients, one to four years post-injury. In that study, another unique set of three subscales were used, which were concluded to be valid. However, in that study, test–retest reliability was not evaluated and responsiveness was not evaluated as a longitudinal measurement, but was calculated as the difference with a pre-injury baseline-group. In this study, the clinimetric measurement properties of the SMFA-NL have been investigated more extensively and in a broader range of trauma patients. Furthermore, the four-subscale configuration of the

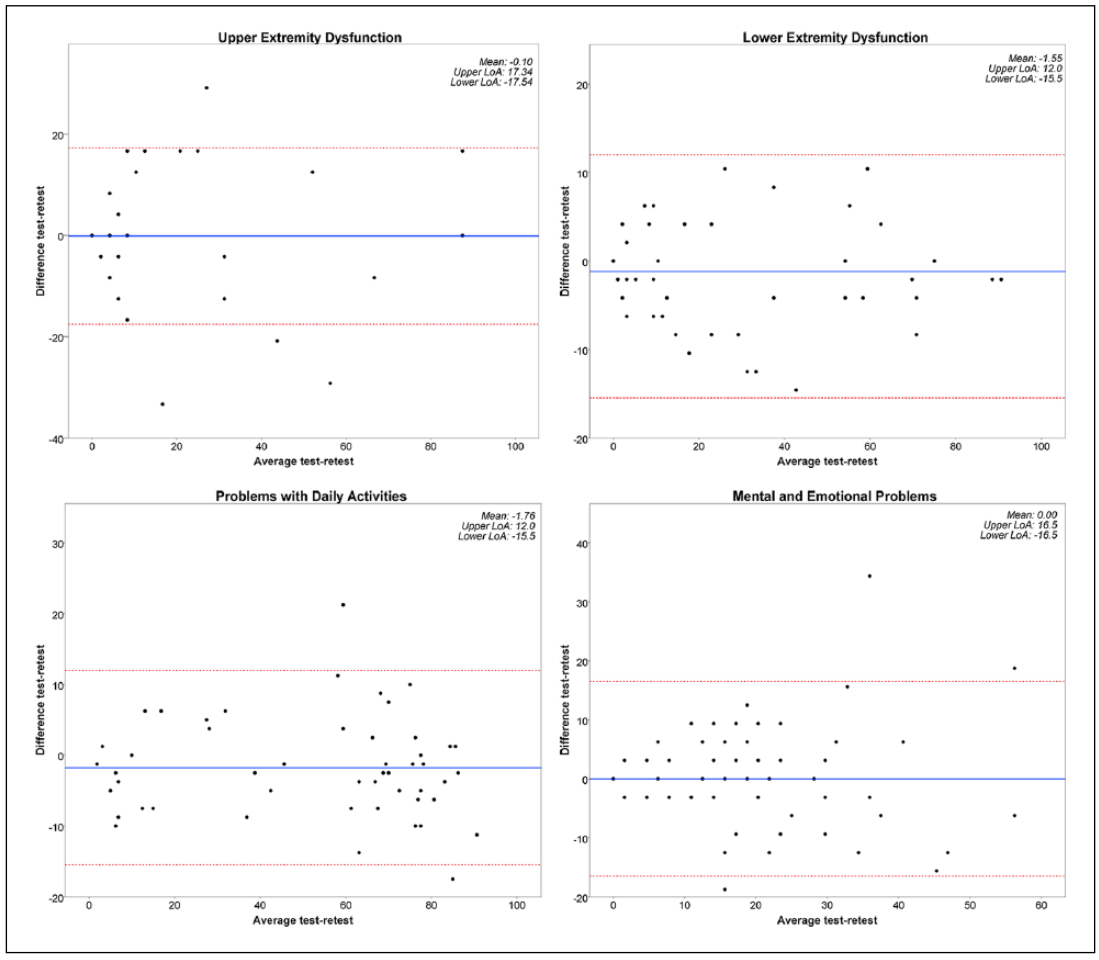

Figure 1. Bland and Altman plots of the test–retest analysis. Bland and Altman plots of the test–retest analysis for all SMFA-NL subscales: Upper Extremity Dysfunction (UED), Lower Extremity Dysfunction (LED), Problems with Daily Activities (PDA) and Mental and Emotional Problems (MEP). Blue line: mean test–retest difference; red dashed lines: limits of agreement. LoA: limit of agreement.

SMFA-NL may be easily applied in day-to-day clinical practice, as there is just one set of subscales for all trauma patients.

The choice of the instruments used for the assessment of construct validity responsiveness was based on the constructs that are evaluated with the subscales of the SMFA-NL. The Disabilities of the Arm Shoulder and Hand (DASH) and Lower Extremity Functional Scale (LEFS) are extremity-specific questionnaires that match the extremity-specific subscales of the SMFA-NL. The Health Utilities Index 3 and the EQ-5D are complementary generic instruments that aim to cover the entire spectrum of disease and functional limitations, including constructs such as daily activities and mental and emotional problems. In a guideline aimed at assessing health status after trauma, it has been advised to use both the Health Utilities Index 3 and EQ-5D.[14]

The floor effect observed for the *Upper Extremity Dysfunction* subscale was mainly caused by patients with a relatively mild injury of the upper extremity.

Table 5. SMFA-NL scores of the six weeks to six months interval.

	$Mean_{6w}$ (SD)	$Mean_{6m}$ (SD)	Mean diff (SD)
Upper Extremity Dysfunction ($n = 159$)	13.2 (20.4)	6.2 (11.8)	7.0 (13.8)
Lower Extremity Dysfunction ($n = 151$)	32.3 (24.7)	17.0 (17.7)	15.3 (19.2)
Problems with Daily Activities ($n = 152$)	53.1 (24.3)	27.9 (21.1)	25.1 (20.9)
Mental and Emotional Problems ($n = 160$)	24.2 (14.4)	20.1 (14.3)	4.1 (12.5)

$Mean_{6w}$: mean of the six weeks post-injury measurement; $Mean_{6m}$: mean of the six months post-injury measurement; mean diff: mean difference of the six weeks and six months post-injury measurements; SMFA-NL: Short Musculoskeletal Function Assessment.

This may indicate that the subscale lacks sensitivity to detect some upper extremity functional problems. Alternatively, these patients may have already recovered after six weeks. Floor effects pose largest problems in longitudinal analyses, as patients cannot show any further improvement in score even if they do experience clinical improvement.[49] Therefore, the use of the *Upper Extremity Dysfunction* subscale may be limited, especially in patients with a relatively mild injury of the upper extremity. Similar to the findings in this study, floor effects have been reported in the development study of the SMFA and in studies of the SMFA that evaluated clinimetric properties of an upper extremity subscale.[1,9,50,51] The addition of items with a higher difficulty may resolve significant floor effects, yet modification of the SMFA-NL was beyond the scope of this study.

The systematic bias of the *Lower Extremity Dysfunction* subscale was considered small and may have been caused by subclinical recovery of the lower extremity. Systematic bias of the SMFA had only been investigated in one study. Reininga et al.[9] reported a small but irrelevant systematic bias for the Bother Index of 2 points. We considered the systematic bias to have had a small influence on the reliability of the *Lower Extremity Dysfunction Subscale*, as the bias was smaller than the measurement error and may be easily controlled for when needed.[52]

The smallest detectable change is an important benchmark to interpret changes in scores. It indicates the point from which a change can be considered a true change and not due to measurement error.[44,53] The smallest detectable change values of the SMFA-NL at group level were considered small. However, at the individual level, the smallest detectable change values of the *Upper Extremity Dysfunction* and *Mental and Emotional Problems* subscales may limit the ability to assess early clinical changes.

Two studies have reported smallest detectable change values of the SMFA.[9,54] Pinsker et al. reported a smallest detectable change of 9.6 points on the function index. This is a much lower smallest detectable change value compared to our findings; however, that was in patients with clinically stable end-stage ankle arthritis, which is not representative of the sample of this study.[54] In a sample of Dutch patients with various musculoskeletal disorders, Reininga et al.[9] reported standard error of measurement values from which smallest detectable change values could be calculated. Despite a slightly higher reliability (intraclass correlation coefficients range: 0.91–0.96) than in this study, the smallest detectable changes of Reininga et al.[9] were larger than in this study (range: 23.3–31.3 points). Reininga et al.[9] studied a more heterogeneous sample, which may have led to higher reliability statistics, while not affecting absolute measurement error. In addition, no external criterion was applied to identify patients who had not changed, which may have increased measurement error. Although the measurement error in this study (expressed as smallest detectable change) was smaller than in the study by Reininga et al., the interpretation of a change in score requires an additional benchmark: the minimal important change.[53] The minimal important change reflects which change in score is a meaningful change to patients. However, a minimal important change is currently not known for the SMFA.[3]

The two-week test–retest interval may be considered a limitation of this study, as it carries the risk of recall bias. However, a two-week test–retest interval is generally considered a safe margin to avoid significant recall bias, but short enough to

avoid clinical improvement.[44] Additional Global Rating of Effect questions were used to exclude that patients experienced clinical change, although these questions may not capture subclinical change. Second, although the sample size was considered adequate, due to the longitudinal study design some patients were lost to follow-up for the eight weeks and six months measurements. This may have induced selection bias. Third, the clinical usability of the Upper Extremity Dysfunction subscale may be limited, especially in patients with a relatively mild injury of the upper extremity.

One of the strengths of this study was that this was the first study in which the responsiveness of the SMFA has been evaluated using the COSMIN guidelines, in which hypotheses-testing is recommended.[41] Other studies have reported standardized response means, which is an effect size-based measure that does not relate to the validity of the measured change.[1,4,5,50,51]

The SMFA-NL may be used in clinical practice as an overall evaluation of physical function at one moment, or as an instrument to assess change in physical function over time. In research the SMFA-NL may be used whenever the researcher is interested in the functional status or functional recovery of an injured patient, for example t in clinical trials in which conservative and surgical treatment of fractures are compared. To improve interpretability of the SMFA-NL, future research may be dedicated to assess which change in score is important to a patient and which difference in score between groups of patients may be considered relevant.

Clinical messages

- The Short Musculoskeletal Function Assessment may be used to assess physical functioning at a single moment in patients who sustained trauma.
- The Short Musculoskeletal Function Assessment may be used to measure change in physical function over time in patients who sustained trauma.
- Floor and ceiling effects may limit the usefulness of the Upper Extremity Dysfunction subscale in longitudinal analyses.

Acknowledgements

All authors have read and agreed the Statement for Authors. M.W.d.G. contributed to data collection, data analysis, data interpretation and drafting the work; I.H.F.R. contributed to data collection, data interpretation and critical revision of the work; K.W.W. and E.H. contributed to critical revision of the work; M.E.M. is the guarantor and contributed to conceptual design of the work, data interpretation and critical revision of the work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Max Willem de Graaf https://orcid.org/0000-0002-2588-5154

Supplemental material

Supplemental material for this article is available online.

References

1. Swiontkowski MF, Engelberg R, Martin DP, et al. Short Musculoskeletal Function Assessment questionnaire: validity, reliability, and responsiveness. *J Bone Joint Surg Am* 1999; 81(9): 1245–1260.
2. Barei DP, Agel J and Swiontkowski MF. Current utilization, interpretation, and recommendations: the Musculoskeletal Function Assessments (MFA/SMFA). *J Orthop Trauma* 2007; 21(10): 738–742.
3. Bouffard J, Bertrand-Charette M and Roy J. Psychometric properties of the Musculoskeletal Function Assessment and the Short Musculoskeletal Function Assessment: a systematic review. *Clin Rehabil* 2016; 30(4): 393–409.
4. Guevara CJ, Cook C, Pietrobon R, et al. Validation of a Spanish version of the Short Musculoskeletal Function Assessment questionnaire (SMFA). *J Orthop Trauma* 2006; 20(9): 623–629; discussion 629–630; author reply 630.
5. Taylor MK, Pietrobon R, Menezes A, et al. Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Short Musculoskeletal Function Assessment questionnaire: the SMFA-BR. *J Bone Joint Surg Am* 2005; 87(4): 788–794.

6. Ponzer S, Skoog A and Bergstrom G. The Short Musculoskeletal Function Assessment questionnaire (SMFA): cross-cultural adaptation, validity, reliability and responsiveness of the Swedish SMFA (SMFA-Swe). *Acta Orthop Scand* 2003; 74(6): 756–763.
7. Lindahl M, Andersen S, Joergensen A, et al. Cross-cultural adaptation and validation of the Danish version of the Short Musculoskeletal Function Assessment questionnaire (SMFA). *Qual Life Res* 2018; 27(1): 267–271.
8. Wang Y, He Z, Lei L, et al. Reliability and validity of the Chinese version of the Short Musculoskeletal Function Assessment questionnaire in patients with skeletal muscle injury of the upper or lower extremities. *BMC Musculoskelet Disord* 2015; 16: 161.
9. Reininga IH, El Moumni M, Bulstra SK, et al. Cross-cultural adaptation of the Dutch Short Musculoskeletal Function Assessment questionnaire (SMFA-NL): internal consistency, validity, repeatability and responsiveness. *Injury* 2012; 43(6): 726–733.
10. Martin DP, Engelberg R, Agel J, et al. Development of a musculoskeletal extremity health status instrument: the Musculoskeletal Function Assessment instrument. *J Orthop Res* 1996; 14(2): 173–181.
11. Engelberg R, Martin DP, Agel J, et al. Musculoskeletal function assessment: reference values for patient and non-patient samples. *J Orthop Res* 1999; 17(1): 101–109.
12. De Graaf MW, Reininga IHF, Wendt KW, et al. Structural validity of the Short Musculoskeletal Function Assessment in patients with injuries. *Phys Ther* 2018; 98(11): 955–967.
13. Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUIÂ®): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003; 1: 54.
14. Van Beeck EF, Larsen CF, Lyons RA, et al. Guidelines for the conduction of follow-up studies measuring injury-related disability. *J Trauma* 2007; 62(2): 534–550.
15. Polinder S, Haagsma JA, Bonsel G, et al. The measurement of long-term health-related quality of life after injury: comparison of EQ-5D and the health utilities index. *Inj Prev* 2010; 16(3): 147–153.
16. Jones CA, Pohar S, Feeny D, et al. Longitudinal construct validity of the Health Utilities Indices mark 2 and mark 3 in hip fracture. *Qual Life Res* 2014; 23(3): 805–813.
17. Slobogean GP, Noonan VK and O'Brien PJ. The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and short form-6D outcome instruments in patients with proximal humeral fractures. *J Shoulder Elbow Surg* 2010; 19(3): 342–348.
18. Blanchard C, Feeny D, Mahon JL, et al. Is the Health Utilities Index valid in total hip arthroplasty patients? *Qual Life Res* 2004; 13(2): 339–348.
19. Jones CA, Feeny D and Eng K. Test-retest reliability of Health Utilities Index scores: evidence from hip fracture. *Int J Technol Assess Health Care* 2005; 21(3): 393–398.
20. Rabin R, Oemar M and Oppe M. *EQ-5D-3L user guide: basic information on how to use the EQ-5D-3L instrument*. Rotterdam: EuroQol Group, 2011.
21. Lamers LM, Stalmeier PF, Krabbe PF, et al. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Making* 2006; 26(2): 173–181.
22. Lamers LM, Stalmeier PF, McDonnell J, et al. Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D tarief. *Ned Tijdschr Genees* 2005; 149: 1574–1578.
23. Derrett S, Black J and Herbison GP. Outcome after injury—a systematic literature search of studies using the EQ-5D. *J Trauma* 2009; 67(4): 883–890.
24. Hung MC, Lu WS, Chen SS, et al. Validation of the EQ-5D in patients with traumatic limb injury. *J Occup Rehabil* 2015; 25(2): 387–393.
25. Grobet C, Marks M, Tecklenburg L, et al. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review. *Arch Orthop Trauma Surg* 2018; 138(7): 953–961.
26. EuroQol Group. Self-complete versions on paper, https://euroqol.org/eq-5d-instruments/eq-5d-3l-available-modes-of-administration/self-complete-on-paper/ (2018, accessed 25 November 2018).
27. Hudak PL, Amadio PC and Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. *Am J Ind Med* 1996; 29(6): 602–608.
28. Institute of Work and Health. Disabilities of the Arm Shoulder and Hand: translations of the DASH questionnaire, http://dash.iwh.on.ca/available-translations (accessed September 2015).
29. Westphal T, Piatek S, Schubert S, et al. Reliability and validity of the upper limb DASH questionnaire in patients with distal radius fractures. *Z Orthop Ihre Grenzgeb* 2002; 140(4): 447–451.
30. Forward DP, Sithole JS and Davis TR. The internal consistency and validity of the patient evaluation measure for outcomes assessment in distal radius fractures. *J Hand Surg Eur Vol* 2007; 32(3): 262–267.
31. Veehof MM, Sleegers EJ, van Veldhoven NH, et al. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther* 2002; 15(4): 347–354.
32. Jones CB, Sietsema DL and Williams DK. Locked plating of proximal humeral fractures: is function affected by age, time, and fracture patterns? *Clin Orthop Relat Res* 2011; 469(12): 3307–3316.
33. Binkley JM, Stratford PW, Lott SA, et al. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther* 1999; 79(4): 371–383.
34. Pan SL, Liang HW, Hou WH, et al. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. *Injury* 2014; 45(11): 1759–1763.
35. Metsavaht L, Leporace G, Riberto M, et al. Translation and cross-cultural adaptation of the lower extremity functional scale into a Brazilian Portuguese version and

validation on patients with knee injuries. *J Orthop Sports Phys Ther* 2012; 42(11): 932–939.
36. Lin CW, Moseley AM, Refshauge KM, et al. The Lower Extremity Functional Scale has good clinimetric properties in people with ankle fracture. *Phys Ther* 2009; 89(6): 580–588.
37. Haitz K, Shultz R, Hodgins M, et al. Test-retest and inter-rater reliability of the functional lower extremity evaluation. *J Orthop Sports Phys Ther* 2014; 44(12): 947–954.
38. Hoogeboom TJ, De Bie RA, Den Broeder AA, et al. The Dutch Lower Extremity Functional Scale was highly reliable, valid and responsive in individuals with hip/knee osteoarthritis: a validation study. *BMC Musculoskelet Disord* 2012; 13: 117.
39. Van der Laan K, Tappan R and Tseng E. Rehab measures: Numeric Pain Rating Scale, http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=891 (2013, accessed December 2016).
40. Karcioglu O, Topacoglu H, Dikme O, et al. A systematic review of the pain scales in adults: which to use? *Am J Emerg Med* 2018; 36(4): 707–714.
41. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; 10: 22.
42. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63(7): 737–745.
43. Acute Zorg Netwerk Noord Nederland. Dutch National Trauma registry, https://www.acutezorgnetwerk.nl/trauma-zorgketen/onderzoek/ (2017).
44. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60(1): 34–42.
45. De Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; 59(10): 1033–1039.
46. Beckerman H, Roebroeck ME, Lankhorst GJ, et al. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 2001; 10(7): 571–578.
47. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476): 307–310.
48. McHorney CA and Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995; 4(4): 293–307.
49. De Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in medicine: a practical guide*. New York: Cambridge University Press, 2011.
50. Van Delft-Schreurs CC, Van Son MA, De Jongh MA, et al. Psychometric properties of the Dutch Short Musculoskeletal Function Assessment (SMFA) questionnaire in severely injured patients. *Injury* 2016; 47(9): 2034–2040.
51. Van Son MA, Den Oudsten BL, Roukema JA, et al. Psychometric properties of the Dutch Short Musculoskeletal Function Assessment (SMFA) questionnaire in patients with a fracture of the upper or lower extremity. *Qual Life Res* 2014; 23: 917–926.
52. Bland JM and Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8(2): 135–160.
53. Terwee CB, Roorda LD, Knol DL, et al. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009; 62(10): 1062–1067.
54. Pinsker E, Inrig T, Daniels TR, et al. Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. *Foot Ankle Int* 2015; 36(6): 617–625.